Climate Predictions with imperfect models

David Sexton

UKCIPnext Planning Retreat, 6-7 June, 2005

Bayesian prediction

- Aim is to construct joint probability distribution $p(X, m_h, m_f, y, o, d)$ of all uncertain objects in problem.
    - Input parameters (X)
    - Historical Model output (m_h)
    - Model prediction (m_f)
    - True climate (y_h,y_f)
    - Observations (o)
    - Model imperfections (d)
- It measures how all objects are related in a probabilistic sense

Best-input assumption

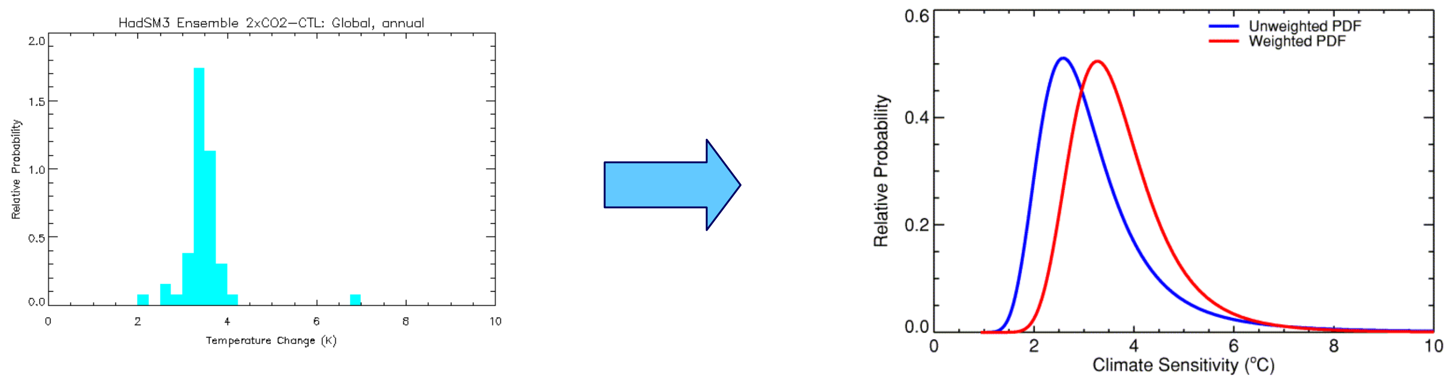
- Start with a perturbed physics ensemble
- Hypothesise that there is a set of input parameters, x*, that provide the best climate model
- But acknowledge that this best model is imperfect and that there is a discrepancy, d, compared to real climate
- We only know the probability that each point in parameter space is the best-input model. But that means we need a model at every part of parameter space…

Emulators and priors

Emulators are statistical models, trained on ensemble runs, designed to predict model output at untried parameter combinations
 (a t-distribution at each sampled point)

Prior distribution $p(m_f)$ – pdf prediction before any observations used

Monte Carlo sampling of parameters combined with an emulator (combining lots of t-distributions) produces prior pdf (blue line).



Comparing models with observations

- Use likelihood function i.e. skill of model is likelihood of model data given some observations

$$\log L_o(\mathbf{m}) = -c - \frac{n}{2}\log|\mathbf{V}| - \frac{1}{2}(\mathbf{m}-\mathbf{o})^T\mathbf{V}^{-1}(\mathbf{m}-\mathbf{o})$$

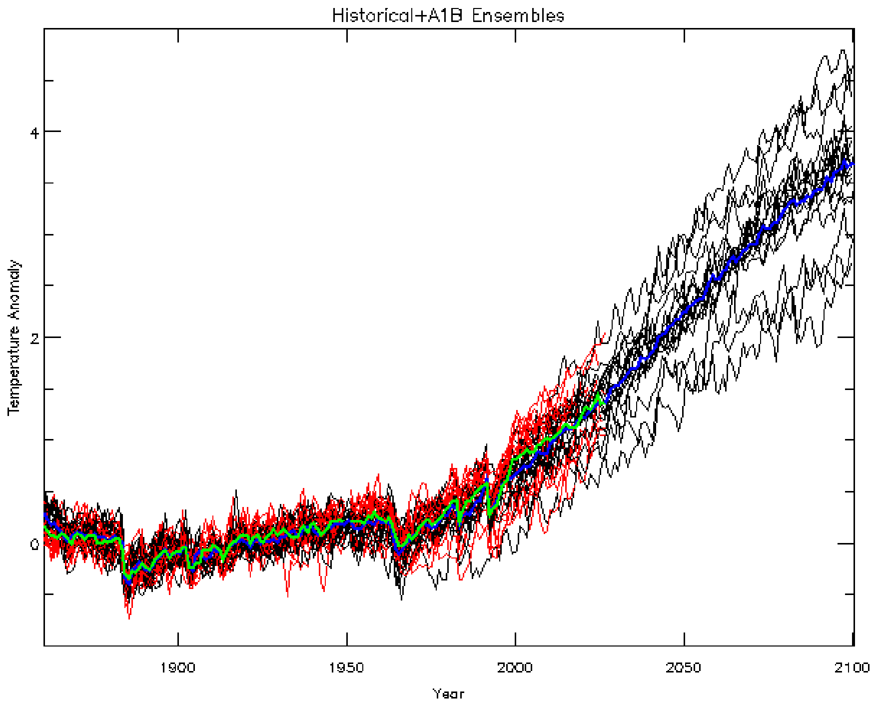
V = observational uncertainty +internal variability+discrepancy

Likelihood used to weight Monte Carlo ensemble members

Estimating discrepancy

- Use multimodel ensemble
- Define discrepancy by some unknown hyperparameters, S.
- For each multimodel ensemble member, find best combination of x* and S that maximises likelihood
- S represents distance between multimodel ensemble member and QUMP i.e. effect of processes not explored by QUMP.
- r.m.s S over multimodel ensemble used to estimate discrepancy

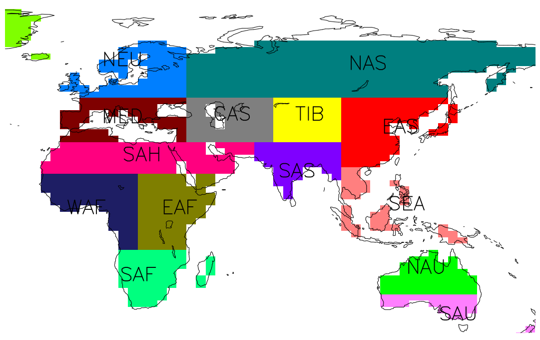
Historical and A1B Scenario

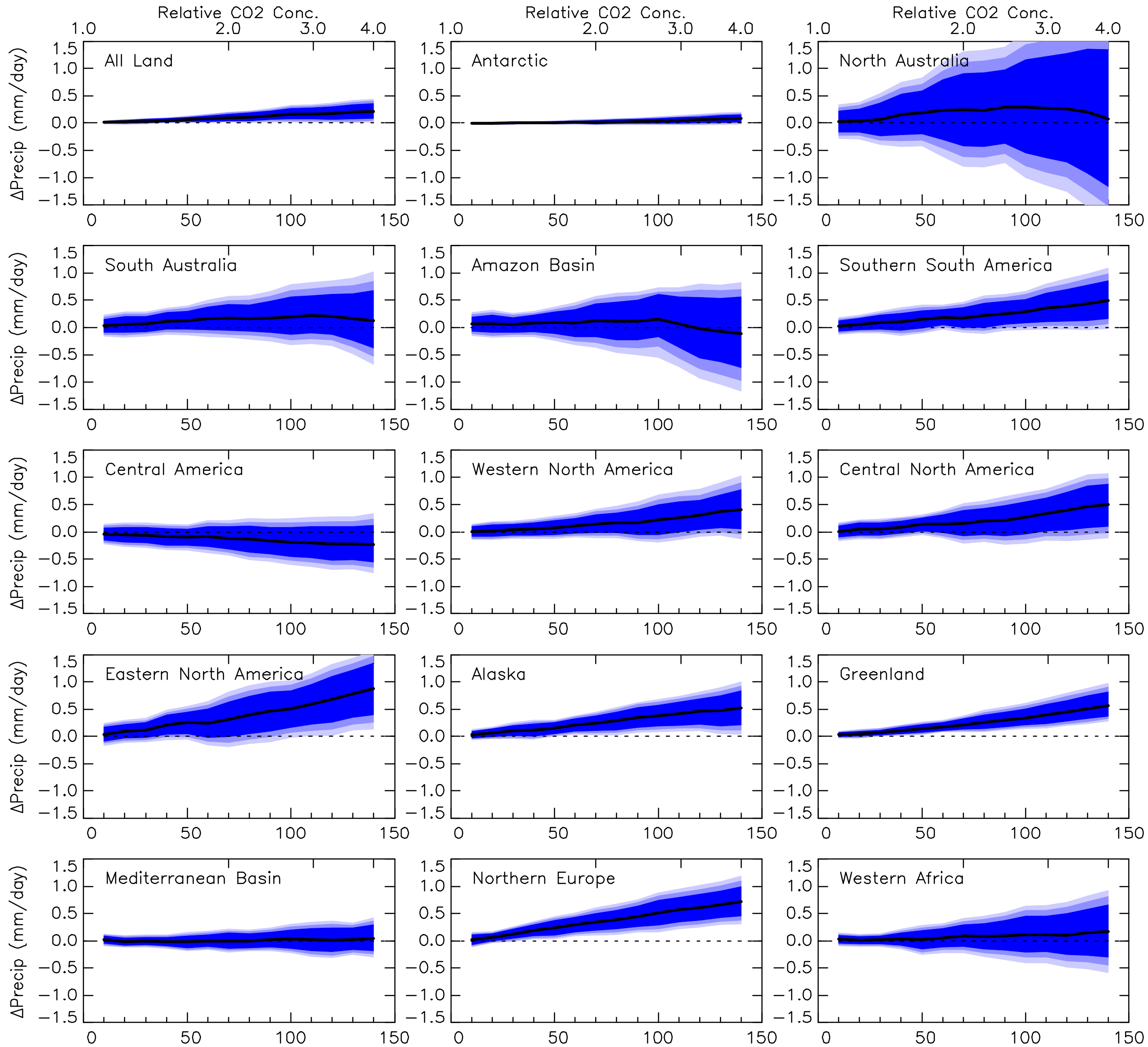


P(x,t)=EBM_global_T(t; slab sensitivity) * slab pattern P(x)

Ben Booth, Glen Harris

Time-scaling Approach





Partial prior predictive distributions

Harris, et al., submitted

Reducing uncertainty

- Improve observational uncertainties
- Improve model i.e. reduce discrepancy
- Run larger ensembles
- Use more observational constraints independent of the ones used already
- Remove pattern scaling and downscaling steps
- Remove assumptions about linking sub-modules

Discrepancy

- Avoids observations over-constraining the pdfs.
    - Avoids case where two sets of observations have constrained two pdfs that seem to contradict each other i.e. don't overlap much.
    - Avoids contradictions from subsequent analyses when some observations have been allowed to constrain the problem too strongly.
- Provides a means of accounting for model quality
    - Model improvements can subsequently be tracked
    - Constraint of observations gradually improve as model improves rather than jumping from "unusable" to "usable".
    - Better models given more weight – physics matters!
- Can possibly be estimated from other climate centre's models and therefore allow for structural uncertainty.